TRƯỜNG ĐẠI HỌC XÂY DỰNG HÀ NỘI

KHOA CÔNG NGHỆ THÔNG TIN

BỘ MÔN: THỊ GIÁC MÁY TÍNH

ĐỀ TÀI:

NHẬN DẠNG CỬ CHỈ TAY

NHÓM 05

Giảng viên hướng dẫn: Thái Thị Nguyệt

Lớp: 65CS1

Sinh viên thực hiện: Trần Quang Huy – 99265

Hà Nội, 6/2023

TRƯỜNG ĐẠI HỌC XÂY DỰNG HÀ NỘI

KHOA CÔNG NGHỆ THÔNG TIN

BỘ MÔN: THỊ GIÁC MÁY TÍNH

ĐỀ TÀI:

NHẬN DẠNG CỬ CHỈ TAY

NHÓM 05

Giảng viên hướng dẫn: Thái Thị Nguyệt

Lớp: 65CS1

Sinh viên thực hiện: Trần Quang Huy – 99265

Hà Nội, 6/2023

Mục lục

1. Lý thuyết

1.1 Giới thiệu

Thị giác máy tính (Computer Vision) là một trong những lĩnh vực hot nhất của khoa học máy tính và nghiên cứu trí tuệ nhân tạo. Dù chúng vẫn chưa thể cạnh tranh với sức mạnh thị giác của mắt người, đã có rất nhiều ứng dụng hữu ích được tạo ra khai thác tiềm năng của chúng.

Một trong những thách thức mà các nhà khoa học máy tính phải vật lộn từ những năm 1950s là tạo ra những cỗ máy có thể hiểu được hình ảnh và video như con người. Lĩnh vực *thị giác máy tính* từ đó đã trở thành một trong những lĩnh vực nghiên cứu hot nhất về khoa học máy tính và trí tuệ nhân tạo.

Nhiều thập kỷ sau, chúng ta đã đạt được tiến bộ lớn trong việc tạo ra các phần mềm có thể hiểu và mô tả nội dung của dữ liệu một cách trực quan. Nhưng chúng ta cũng đã nhận ra rằng cần phải đi xa đến mức nào trước khi có thể hiểu và tái tạo một trong những chức năng cơ bản của bộ não con người.

1.2 Đề tài thực hiện

Đề tài "Nhận diện ký hiệu tay” sử dụng kỹ thuật Convolutional Neural Network (CNN) là một lĩnh vực nghiên cứu trong lĩnh vực Trí tuệ nhân tạo và Thị giác máy tính. Nó tập trung vào việc phát hiện và phân loại các ký hiệu tay được tạo ra bởi người dùng để truyền tải thông điệp hoặc tương tác với hệ thống máy tính.

CNN là một mạng neural network đặc biệt được thiết kế để xử lý dữ liệu có cấu trúc ruột, như ảnh và video. Kiến trúc của CNN chứa các lớp tích chập (convolutional layers) và các lớp tổng hợp (pooling layers), giúp nắm bắt được các đặc trưng cục bộ của hình ảnh. CNN đã chứng tỏ khả năng xuất sắc trong việc xử lý ảnh và phân loại các đối tượng trong ảnh.

Ứng dụng của đề tài này rất đa dạng, từ việc điều khiển thiết bị thông qua các ký hiệu tay, hỗ trợ người khuyết tật trong giao tiếp, cho đến các ứng dụng trong lĩnh vực thị giác máy tính và thực tế ảo. Với khả năng học và nhận diện các ký hiệu tay, các hệ thống dựa trên CNN có thể tạo ra một trải nghiệm tương tác tự nhiên và hiệu quả giữa con người và máy tính.

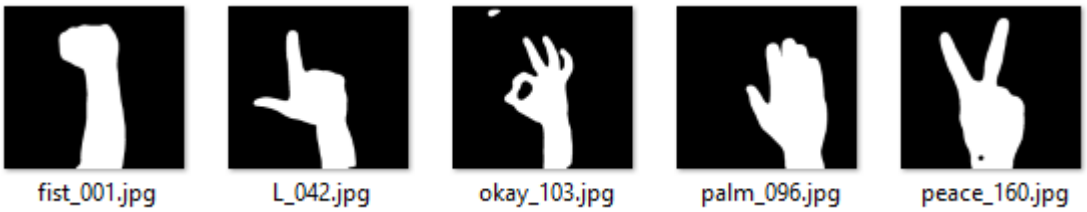
2. Thực hiện

Để triển khai đề tài này, quy trình thường bao gồm các bước sau:

1. Xây dựng tập dữ liệu: Thu thập và chuẩn bị dữ liệu với các hình ảnh chứa các ký hiệu tay. Tập dữ liệu này phải đủ lớn và đa dạng để mô hình có thể học các biểu đạt khác nhau của ký hiệu tay.
2. Tiền xử lý dữ liệu: Chuẩn hóa và tiền xử lý dữ liệu để đảm bảo tính nhất quán và khả năng so sánh giữa các hình ảnh. Điều này bao gồm việc chuyển đổi hình ảnh thành định dạng thích hợp (đen trắng), thay đổi kích thước hình ảnh và chuẩn hóa các giá trị pixel.
3. Xây dựng mô hình CNN: Xây dựng một mô hình CNN với các lớp tích chập, lớp tổng hợp và các lớp kết nối đầy đủ (fully connected layers) để học các đặc trưng và phân loại ký hiệu tay. Các tham số của mô hình sẽ được tinh chỉnh thông qua quá trình huấn luyện trên tập dữ liệu.
4. Huấn luyện mô hình: Huấn luyện mô hình CNN trên tập dữ liệu đã chuẩn bị. Quá trình này sẽ điều chỉnh các trọng số và các tham số của mô hình để tối ưu hóa hiệu suất phân loại.
5. Đánh giá và kiểm tra: Đánh giá mô hình trên tập dữ liệu kiểm tra độc lập để đo lường hiệu suất và độ chính xác của mô hình. Điều này giúp xác định khả năng tổng quát hóa của mô hình trong việc phân loại các ký hiệu tay mới.

2.1 Train model (huấn luyện mô hình)

- Để có thể thực hiện train model, cần có tập dữ liệu là các ảnh đầu vào là các ảnh bàn tay đang thực hiện các ký hiệu, các ảnh đã được gán nhãn đầy đủ.



Ảnh1: Các ảnh đã được gán nhãn đầy đủ bằng tên file.

- Sau đó chuyển các ảnh về size 224x224.
- Dữ liệu ảnh được chuyển đổi sang định dạng numpy và chuẩn hóa về khoảng [0, 1]. Nhãn cử chỉ được chuyển đổi sang dạng one-hot encoding.
- Phân chia dữ liệu thành tập train và test với tỉ lệ là 80/20.
- Khởi tạo mô hình với mạng CNN đã được thêm một vài layer Dense (Fully Connect) và cuối cùng là một lớp softmax để dự đoán kết quả đầu ra, ở đây là 5 ký hiệu tay.
- Thực hiện train mạng với khoảng hơn 2000 ảnh train và gần 300 ảnh test.
- Sau khi train xong kết quả sẽ được lưu lại.

```
Epoch 42/50
2198/2198 [==============================] - 8s 3ms/step - loss: 5.7508e-05 - accuracy: 1.0000 - val_loss: 0.0858 - val_accuracy: 0.9891
Epoch 43/50
2198/2198 [==============================] - 8s 3ms/step - loss: 9.2834e-05 - accuracy: 1.0000 - val_loss: 0.0786 - val_accuracy: 0.9891
Epoch 44/50
2198/2198 [==============================] - 8s 3ms/step - loss: 1.9233e-05 - accuracy: 1.0000 - val_loss: 0.0791 - val_accuracy: 0.9909
Epoch 45/50
2198/2198 [==============================] - 8s 3ms/step - loss: 1.1290e-05 - accuracy: 1.0000 - val_loss: 0.0799 - val_accuracy: 0.9909
Epoch 46/50
2198/2198 [==============================] - 8s 3ms/step - loss: 2.1392e-05 - accuracy: 1.0000 - val_loss: 0.0815 - val_accuracy: 0.9909
Epoch 47/50
2198/2198 [==============================] - 8s 3ms/step - loss: 3.1780e-04 - accuracy: 1.0000 - val_loss: 0.1034 - val_accuracy: 0.9873
Epoch 48/50
2198/2198 [==============================] - 8s 3ms/step - loss: 5.5440e-05 - accuracy: 1.0000 - val_loss: 0.0885 - val_accuracy: 0.9909
Epoch 49/50
2198/2198 [==============================] - 8s 3ms/step - loss: 2.5760e-05 - accuracy: 1.0000 - val_loss: 0.0846 - val_accuracy: 0.9891
Epoch 50/50
2198/2198 [==============================] - 8s 3ms/step - loss: 1.1906e-05 - accuracy: 1.0000 - val_loss: 0.0828 - val_accuracy: 0.9891
```

Ảnh 2: Quá trình train dữ liệu

2.2 Detection (phát hiện)

- Thực hiện mở Webcam và khoanh 1 vùng sẽ phát hiện bàn tay ở trong vùng đó. Gọi là detection region.

- Người dùng đưa tay vào trong vùng detection và tạo thành các hình ký tự.
- Thực hiện xoá nền trong ảnh thu được.

Ảnh 3: Sau khi tách nền ra khỏi ảnh gốc

- Mô hình dự đoán trên ảnh và trả về một mảng các xác suất dự đoán cho các lớp.

- Model sẽ hiển thị ký tự ra màn hình.

Ảnh 4: Kết quả dự đoán

3. Kết luận

3.1 Đánh giá

Mô hình nhận dạng ký hiệu tay hoạt động và cho ra một kết quả tương đối tốt, tuy nhiên điểm khó của mô hình CNN bên trên là phải tinh chỉnh các tham số như số lượng bộ lọc, số lương nơ-ron trong các lớp Dense, tốc độ học, thuật toán tối ưu hoá, số lượng epoch, kích thước batch.

Để tối ưu hoá mô hình, ta phải thử nghiệm và điều chỉnh các yếu tố trên để có thể tăng khả năng học và giảm overfitting, cũng như thực hiện tăng cường dữ liệu và thực hiện đánh giá hiệu suát của mô hình trên tập dữ liệu test.